



**QUEEN'S
UNIVERSITY
BELFAST**

Comparison of methods used to calculate typical threshold values for potentially toxic elements in soil

McIlwaine, R., Cox, S., Doherty, R., Palmer, S., Offerdinger, U., & McKinley, J. M. (2014). Comparison of methods used to calculate typical threshold values for potentially toxic elements in soil. *Environmental Geochemistry and Health*, 36(5), 953-971 . <https://doi.org/10.1007/s10653-014-9611-x>

Published in:

Environmental Geochemistry and Health

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

Copyright 2014 Springer Science+Business Media Dordrecht.

This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Comparison of methods used to calculate typical threshold values for potentially toxic elements in soil

Rebekka McIlwaine^{1*}, Siobhan F. Cox¹, Rory Doherty¹, Sherry Palmer¹, Ulrich Ofterdinger¹, Jennifer M. McKinley²

¹Environmental Engineering Research Centre, School of Planning, Architecture and Civil Engineering, Queen's University Belfast, Belfast, BT9 5AG, UK

²School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, Belfast, BT9 5AG, UK

*Address correspondence to Rebekka McIlwaine, Environmental Engineering Research Centre, School of Planning, Architecture and Civil Engineering, Queen's University Belfast, Belfast, BT9 5AG, E-mail:rmcilwaine05@qub.ac.uk

1 Abstract

The environmental quality of land can be assessed by calculating relevant threshold values which differentiate between concentrations of elements resulting from geogenic and diffuse anthropogenic sources and concentrations generated by point sources of elements. A simple process allowing the calculation of these typical threshold values (TTVs) was applied across a region of highly complex geology (Northern Ireland) to six elements of interest; arsenic, chromium, copper, lead, nickel and vanadium. Three methods for identifying domains (areas where a readily identifiable factor can be shown to control the concentration of an element), were used: k-means cluster analysis, boxplots and empirical cumulative distribution functions (ECDF). The ECDF method was most efficient at determining areas of both elevated and reduced concentrations and was used to identify domains in this investigation. Two statistical methods for calculating Normal Background Concentrations (NBCs) and Upper Limits of Geochemical Baselines Variations (ULBLs), currently used in conjunction with legislative regimes in the UK and Finland respectively, were applied within each domain. The NBC methodology was constructed to run within a specific legislative framework, and its use on this soil geochemical data set was influenced by the presence of skewed distributions and outliers. In contrast, the ULBL methodology was found to calculate more appropriate TTVs that were generally more conservative than the NBCs. TTVs indicate what a “typical” concentration of an element would be within a defined geographical area and should be considered alongside the risk that each of the elements pose in these areas to determine potential risk to receptors.

2 Suggested Keywords

Background; contaminated land; domain identification; threshold; NBC; ULBL

3 Introduction

Geochemical surveys are carried out for various different reasons. Initially, they were used to define the extent of mineralised areas in prospectivity studies (Hawkes and Webb 1962) and often urban areas would have been avoided in these surveys (Johnson and Ander 2008). However, with developments in the understanding of the effects potentially toxic elements (PTEs) have on the environment and human health, geochemical surveys are increasingly being used in investigations to determine land quality and contamination (Salminen and Tarvainen 1997). A fundamental aim of geochemical surveys is often to define PTE concentrations that provide relevant thresholds within spatial element distributions. Originally used as a prospecting tool (Sinclair 1974), threshold values are increasingly employed as a method by which to discriminate “contaminated land” (Rodrigues et al. 2009). In this respect, the threshold is often set to differentiate between concentrations of the element that naturally occur in the soil and concentrations that result from diffuse anthropogenic sources, or even to differentiate between diffuse and point anthropogenic sources. However, there remains little consensus on what the aim of calculating these values is, and how values should be calculated.

Many terms are used in the literature to describe concentrations of elements in the soil, often with conflicting or overlapping definitions. In order to distinguish between geogenic and anthropogenic contamination Matschullat et al. (2000) define the geochemical background as a “relative measure to distinguish between natural element or compound concentrations and anthropogenically influenced concentrations”, which is similar to Hawkes and

Webbs' (1962) definition of background as “the normal abundance of an element in barren earth material”. However British Standards (BS19258) state that the background content of a substance in soil results from both geogenic sources and diffuse source inputs and that the background values should be a “statistical characteristic of the background content” (British Standards 2011) which is therefore similar to Salminen and Tarvainen's (1997) definition of a geochemical baseline as an element's average concentration in the Earth's crust regardless of the source. A discussion by Reimann and Garrett (2005) examines in detail the various terms used to describe these values, including background, threshold, natural background and baseline and the many definitions that exist for these terms.

Salminen and Tarvainen (1997) suggest that baseline values are of “essential importance in environmental legislation” to define limits of PTEs in contaminated land and recent changes in contaminated land legislation in England and Wales have recognised this by stating that “normal levels of contaminants in soil should not be considered to cause land to qualify as contaminated land, unless there is a particular reason to consider otherwise” (Defra 2012). Similarly, a recent Government Decree in Finland on the Assessment of Soil Contamination and Remediation Needs (Ministry of the Environment Finland 2007) requires the input of geochemical baseline concentrations in Finnish soils during the assessment process. An investigation of arsenic concentrations at a site of specific interest in southern Italy, led to the development of a statistical methodology for determining the difference between natural and anthropogenic concentrations of metals and metalloids in soils (APAT-ISS 2006). This methodology was retained by the Italian government as it was considered to be not only applicable to this particular site, but also to all other sites of national interest where the same problem was occurring.

Within the research described in this paper, the term TTV is used to refer to a value which gives a characteristic concentration for an element within a defined geographical area known as a domain. Previous work by Ander et al (2013a) and Ander et al (2013b) has seen the development of a methodology to determine NBCs of contaminants in English soils, supporting the recent changes to the statutory guidance (Defra 2012). Within this methodology, a domain was defined as an area in which a readily distinguishable factor could be identified as controlling the concentration of the element. This approach has been maintained within this investigation, remembering that these areas need to be defined on an element by element basis using initial assessments of the distribution of the elements within the study area. It is important that the methods used to identify domains take all the relevant factors affecting soil element concentrations into account; geogenic factors, diffuse source anthropogenic inputs and point source contamination. In order to be most relevant and useful for environmental legislation, the typical threshold values calculated should define concentrations of PTEs that are typical of the threshold between geogenic and diffuse anthropogenic source contributions to soil and concentrations that are associated with point sources. If point sources of anthropogenic contamination can be identified, they can be more readily assessed to determine if they pose any risk to the surrounding environment. A number of different industries can make use of definite concentrations which achieve this differentiation. In particular, contaminated land professionals can more easily determine sites that possibly require further investigations because the TTVs are exceeded. In addition, the agricultural industry may be interested in depleted concentrations of these elements where they are also considered to be essential to animal and plant life e.g. copper.

Commonly investigated PTEs include arsenic (As), cadmium (Cd), cobalt (Co), copper (Cu), chromium (Cr), iron (Fe), mercury (Hg), manganese (Mn), nickel (Ni), lead (Pb), vanadium (V) and zinc (Zn) (Ajmone-Marsan et al. 2008; Kelepertsis et al. 2006; Palmer et al. 2013; Paterson et al. 2003; Ramos-Miras et al. 2011). Concentrations of PTEs are assessed for a variety of reasons; As, Hg and Pb are examples of elements commonly investigated in urban areas (Chirenje et al. 2003, 2004; Rodrigues et al. 2006; Wong et al. 2006), while Cr, Ni, V and Zn have previously been investigated as geogenically controlled PTEs within Northern Ireland (Cox et al. 2013; Palmer et al. 2013). Previous research has identified concentrations of As, Cd, Cr, Cu, Ni, Pb, V and Zn in Northern Irish soils that exceed relevant Generic Assessment Criteria (GAC)/Soil Guideline Values (SGVs) (Barsby et al. 2012; Martin et al. 2009a; Martin et al. 2009b; Nathanail et al. 2009). Six PTEs have been selected for investigation in this research; As, Cr, Cu, Ni, Pb and V. These elements are expected to be governed by a mixture of geogenic and anthropogenic sources, a necessary factor in order to complete the aims of this study.

The rationale behind this research is to investigate soil geochemical data for Northern Irish soils by 1) using a variety of techniques to identify the principal controls on the spatial variation of the PTEs and determining which technique is most appropriate for the available data set; 2) identifying domains i.e. areas of elevated and reduced PTE concentrations; 3) using previously developed statistical methodologies to calculate TTVs of PTEs within the aforementioned areas and 4) critically comparing the values calculated to determine which statistical method is most appropriate for use in differentiating between diffuse and point source element concentrations. It is worth noting that the TTVs do not assess risk, but instead provide an indication of PTE concentrations that are typical at a site.

4 Materials and Methods

4.1 Study area

Northern Ireland is part of the United Kingdom which sits in the north east of the island of Ireland (Figure 1a) and is home to over 1.8 million people. Despite being less than 14,000 km² in area, the bedrock in Northern Ireland (Figure 1b) ranges from Mesoproterozoic to Palaeogene in age and as a result is said to present an “opportunity to study an almost unparalleled variety of geology in such a small area” (Mitchell 2004). The bedrock is often simplified into a series of Caledonian terranes and part of a Palaeogene igneous province with distinct geological characteristics. The psammites in the northwest of Northern Ireland are of Neoproterozoic age. The south eastern terrane is Lower Palaeozoic in age and also contains younger igneous intrusions. These Palaeogene igneous intrusions consist of three central complexes; the Mourne Mountains, Slieve Gullion and Carlingford. The southwest comprises of a mixture of sandstones, mudstones and limestones, which are mainly Upper Palaeozoic in age with a distinct Lower Palaeozoic inlier. The north east is dominated by a large area of extrusive Palaeogene basalts. In terms of superficial geology, peatlands cover 12% of land area (Davies and Walker 2013) as shown on Figure 1c. Two main urban areas exist within the country; Belfast and Londonderry, with populations of approximately 280,000 and 108,000 respectively (NI Statistics & Research Agency 2013), with other smaller urban centres including towns and villages (Figure 1c).

4.2 Soil geochemical data

The Tellus project, managed by the Geological Survey of Northern Ireland (GSNI), comprised both geophysical and geochemical surveys. The geochemical survey saw the collection of nearly 30,000 soil, stream-sediment and stream-water samples across Northern Ireland between 2004 and 2006. Urban and regional soil samples were collected at densities of 4 per km² and 1 per 2 km² respectively. Two depths were sampled at each location; a shallow sample taken between 5 and 20 cm and a deeper sample taken between 35 and 50cm. The sample taken at each location was a composite of auger flights collected at the four corners and the centre of a 20 by 20m square. Samples were air-dried at the field-base before transport to the sample store where they were oven dried at 30°C for approximately two to three days. The shallow samples were shipped to British Geological Survey (BGS) laboratories in Keyworth, Nottingham for preparation and analysis via x-ray fluorescence (XRF). Sample preparation entailed sieving to a <2mm fraction, from which a sub-sample was produced for milling and pressed pellet production.

A number of quality control methods were employed during the XRF analysis. Two duplicate and two replicate samples were analysed per batch of 100 samples. Three secondary reference materials that were collected in Northern Ireland specifically for the Tellus survey, and one material from BGS's Geochemical Baselines Survey of the Environment (G-BASE) program, were routinely analysed at a rate of two insertions per batch. Certified reference materials were also analysed before and after each batch. Further details of quality control methods are provided by Smyth (2007).

4.3 Domain Identification

Fig. 1 Maps showing **a** location **b** simplified bedrock geology and **c** areas of peat substrate (superficial geology), rural and urban areas across Northern Ireland (Bedrock and superficial geology derived from data provided by GSNI (Crown Copyright))

4.3.1 Known controls over PTEs

In order to calculate TTVs, domains were identified for each element. Domains were selected based on knowledge of the factors shown in Figure 1 which were identified as the main controls over element concentrations in soils.

Studies have shown that the majority of glacial till in Northern Ireland is found within only a few kilometres of its origin suggesting that soils usually reflect the character of the underlying geology (Cruickshank 1997; Jordan 2001). Therefore bedrock geology is expected to provide a strong control over element concentrations in soil. Geochemically, it is likely that the extrusive and intrusive (in particular, the Antrim Basalt formation and Mourne Mountain Complex) igneous rocks of Northern Ireland will be of most interest, as previous studies have shown that they contain reduced and elevated concentrations of a number of elements (Barrat and Nesbitt 1996; Green et al. 2010; Hill et al. 2001; Smith and McAlister 1995; Smyth 2007). Previous studies have demonstrated that soils from the basalt area are more homogeneous in their geochemical content than soils from areas of other rock types, suggesting that the basalts are acting as the soil parent material and the main control over geochemistry in that area (Zhang et al. 2007). A simplified representation of bedrock geology in Northern Ireland derived from GSNI's 1:250000 bedrock geology map is shown in Figure 1b, grouping bedrock types of similar composition and age.

Existing literature suggests that areas of peat substrate within Northern Ireland have a control over the distribution of a number of elements (Palmer et al. 2013). Peat bogs that are fed solely by atmospheric deposition (ombrotrophic), can be used as archives of many types of atmospheric constituents (Shotyk 1996), including contamination in the form of PTEs. Topographically elevated areas of peat are more likely to be affected in this way, as increased precipitation is usually associated with elevation (Goodale et al. 1998). Areas of peat were defined using data derived from the GSNI's 1:250000 superficial geology map and are shown on Figure 1c.

Urban and rural areas were defined using a revised version of the Corine Land Cover 2006 seamless vector data (European Environment Agency 2012). This approach is different to that taken in the NBC methodology, where the Generalised Land Use Database Statistics for England 2005 (Communities and Local Government 2007) were used. The Corine Land Cover data set covers all of Europe and defines 44 land use classes based on the interpretation of satellite images (European Environment Agency 2012). This has been simplified in Figure 1c to show urban and rural areas in Northern Ireland. The majority of the land use classes were easily defined as either rural or urban, with a few others defined on a site by site basis.

In the NBC methodology, metalliferous mineralisation and mining maps, were used to define mineralisation domains throughout the study (Ander et al. 2011). As this information is not available for Northern Ireland, a different approach was taken in using mineral occurrence locations provided by GSNI, alongside relevant literature (Lusty et al. 2009, 2012; Parnell et al. 2000) to aid in mineralised domain identification.

4.3.2 Method used for domain identification

It is important that the methods used to define domains be robust to non-normality and the presence of outliers that are common in geochemical data. Three methods to aid in the domain identification process were compared in this study; k-means cluster analysis (Ander et al. 2011), boxplot mapping (Reimann et al. 2008) and Empirical Cumulative Distribution Function (ECDF) mapping (Reimann 2005). All statistical analysis of data was completed in the R statistical software package (R Core Team 2013), and all geographical analysis and images were completed using ArcMap 10.0 (ESRI 2009).

The k-means cluster method (Figure 2a) was used to define domains by Ander et al. (2011) in the NBC methodology. As k-means cluster analysis is of the partitional variety, the number of clusters must be assigned to the technique at the outset (Jain et al. 1999). The most visually acceptable number of clusters, based on an antecedent visual assessment (Templ et al. 2008), was input into the technique and the data were partitioned into the selected number of clusters by minimising the "average of the squared distances between the observations and their cluster centres" (Reimann et al. 2008). The algorithm constructed by Hartigan and Wong (1979), generally considered to be the most efficient (Ander et al. 2011), was used as the default setting in the R software package (R Core Team 2013). Each data point was classified into a cluster by the technique, allowing the creation of a map of the clusters across Northern Ireland.

Tukey boxplots of the log-transformed data (Figure 2b) were also used to define the classes for producing maps of the data distribution. Assumptions regarding normal distribution of the data appear in the boxplot construction when the whisker values are calculated, as their calculation (box extended by 1.5 times the length

of the box in both directions) assumes data symmetry (Reimann et al. 2008). Log transformations were applied as geochemical data are often strongly right-skewed, and the log-transformation helps the data distribution to approach symmetry, allowing a better visual demonstration of the data when mapped (Reimann et al. 2008). The boxplot was used to split the element concentrations into five classes for mapping; lower extreme values to lower whisker, lower whisker to lower hinge, lower hinge to upper hinge i.e. the box, upper hinge to upper whisker and upper whisker to upper extreme values.

The third method applied to map the distribution of the elements used classes based on the empirical cumulative distribution function (ECDF) (Sinclair 1974). The ECDF graph is a discrete step function which jumps by $1/n$ at each of the n data points. As shown in Figure 2c, the ECDF plots have been constructed using the log-transformed concentrations of the element, in this case nickel, in order to make breaks in the distribution more obvious. Breaks in the distribution are demonstrated through changes of gradient in the graph, and are likely to be caused by the presence of different sub-populations within the data set, with different underlying factors controlling the concentrations of elements in these populations (Díez et al. 2007; Reimann et al. 2008; Reimann et al. 2005). Therefore breaks in the distribution can be used to distinguish mapping class boundaries.

4.3.3 Domain Corroboration

In order to corroborate the results of the domain identification process, a geostatistical approach involving the construction of semi-variograms was used to ensure that the controlling factors over element concentrations in soil were correctly identified. The geostatistics were generated using ArcMap 10.0 (ESRI 2009). A semi-variogram is based on the Theory of Regionalised Variables (Matheron 1965), which permits interpolation of values on a surface by assuming that data points closest to each other spatially will have a greater influence over estimated values than would data points further away from each other. Several important pieces of information can be identified from the semi-variogram:

- Spatial variation at a finer scale than the sample spacing (Deutsch and Journel 1998) and measurement error (Journel and Huijbregts 1978) is represented by the nugget (C_0). Such small scale sources of variance can be an indication that sampling or analytical error is present or, that micro-scale processes are governing geochemistry to a greater degree than was detected by sampling resolutions.
- The spatially correlated variation is represented by the structured component (C_I) (Lloyd 2007).
- The sill (C_x), where the semi-variogram levels off, is the distance at which pairs of data points are no longer spatially dependent upon each other.
- The nugget:sill ratio (C_0/C_x) gives the proportion of random to spatially structured variation at the scale being investigated.
- Ranges of influence (a) can be statistically inferred from the lag distance at which the sill is reached (McKinley et al. 2004), permitting interpretation of specific environmental factors that may be influencing the mapped element of interest.
- Depending upon the nature of fitting measured values to a semi-variogram, multiple spatial structures can be identified. This is of particular interest where investigations into multiple environmental factors that may be controlling the element of interest are required.

- An apparent lack of spatial structure also provides important information, such as giving an indication about the suitability of analytical or sampling methods in accurately detecting total element concentrations which are thoroughly representative of a particular study area.

4.4 Calculation of Typical Threshold Values

In response to the legislative requirements discussed in the introduction, different authors have derived methods by which “background” values can be calculated. The NBC methodology (Ander et al. 2013a; Ander et al. 2013b) aims to provide a mechanism for revised legislation that differentiates between levels of contamination from geogenic and diffuse sources and those from point source contamination. In order to take account of spatial variability, domains were defined for each element by comparing the results of a k-means cluster analysis to a soil parent material model, land use classifications and mineralisation and mining geographical mapping. Within the methodology, it is recommended that the domains are based on at least 30 values. The NBC was then calculated for each domain using a statistical methodology that (1) assesses the skewness of the geochemical data by observing a histogram and calculating the skewness and octile skewness of the distribution. Based on the results of that assessment, the method (2) performs either a log transformation or a box-cox transformation on the data if necessary and then (3) computes percentiles using either parametric, robust or empirical methods depending on the results of the transformation applied. The NBC is then taken to be the upper 95% confidence limit (UCL) of the 95th percentile. A detailed explanation of how the methodology was constructed and how it should be applied is given in Cave et al. (2012).

Jarva et al. (2010) have developed a methodology to allow the calculation of “baselines” in Finland, which in this instance refer to both the “natural geological background concentrations and the diffuse anthropogenic input of substances at regional scale”. As with the above NBC calculations, Finland was divided into a number of geochemical provinces. A key difference between the two methodologies is the consideration of soil type, with baseline values calculated by soil type within geochemical provinces. The ULBL is based on the upper limit of the upper whisker line of the box and whisker plot. A box and whisker plot identifies any values which fall above the upper whisker line as outliers, which may “represent natural concentrations of an element at the sampling site” (Jarva et al. 2010) but are probably not typical of the geochemical province as a whole. Logarithmic transformed data were not used to plot the box and whisker plots, as the untransformed data led to the highest amount of outliers and therefore was felt to give a more conservative value.

A key difference between the NBC and ULBL methodology is the determination of what a “conservative” value is considered to be. The ULBL methodology aims to identify the maximum number of outliers, therefore generating a lower concentration for the ULBL and the possibility that larger areas of land will be identified as exceeding the ULBL. The NBC methodology supports the English contaminated land regime, which aims to identify sites where “if nothing is done, there is a significant possibility of significant harm such as death, disease or serious injury” (Ander et al. 2013a). Therefore, by taking the upper 95% confidence limit of the 95th percentile, the aim seems to be to identify the highest risk sites in order to prioritise further investigation and management of these sites.

Within this research, both the NBC and the ULBL methodologies were applied to the shallow XRF data available for all of Northern Ireland. NBC calculations were carried out using the R scripts prepared by Cave et

al. (2012) while ULBL calculations were undertaken in R using scripts prepared by the authors. Data from shallow soils were selected for analysis so both anthropogenic and geogenic influences on the element concentrations could be determined. XRF was selected as the most appropriate analytical method as it is said to give total results (Ander et al. 2013a).

5 Results and Discussion

5.1 Comparison of domain identification methods

Fig. 2 Domain identification methods completed for Ni concentrations in the shallow soils of Northern Ireland analysed by XRF; **a** completed by a k-means cluster analysis, **b** classes defined by boxplot of log transformed concentrations as shown, and **c** classes defined by ECDF of log transformed concentrations as shown, with inverse distance weighting used to map the results (output cell size of 250m, power of two and a fixed search radius of 1500m)

Figure 2 gives a comparison of the three methods used to map the distribution of elements and therefore identify domains using Ni as an example. The k-means map (Figure 2a) highlights only the basalts which overlie northeast Northern Ireland. Both the boxplot map (Figure 2b) and the ECDF map (Figure 2c) show areas of elevated and reduced concentrations. Both show elevated concentrations over the basalts, with the boxplot method mapping the boundary of the basalts most effectively. Reduced concentrations are more easily identified through the ECDF map, and are obviously correlated to the Mourne Mountain complex of south eastern Northern Ireland. By comparing this image with areas of peat substrate (Figure 1c), an association between areas of peat and reduced concentrations was also identified.

The k-means technique, used in the NBC methodology, produces useful results in the determination of elevated domains; however, it is more commonly used to compare a number of variables and estimate which variables are similar and dissimilar to each other (Romesburg 2004) and the inability of the method to determine domains of reduced concentrations does limit its applicability in practice. In the case of Ni (Figure 2a), the initial assessment demonstrated that 3 clusters would be most appropriate, however a certain amount of prior knowledge regarding the controls over element concentrations is expected in this assessment.

Boxplots allow identification of both elevated and reduced concentrations of the elements, with different sections of the distribution related to separate parts of the boxplot. However, the splits in the distribution are still set at arbitrary values within the dataset, meaning actual controlling factors over the element concentrations could be missed.

The ECDF method was superior in terms of spatially identifying both elevated and depleted concentrations of elements as it retains a great deal of information about the distribution of the element in the mapped output. It clearly delineates areas of both elevated and depleted concentrations allowing controls over the PTE concentrations to be determined. However, the method does require a level of interpretation as the individual inspecting the graphs decides where the gradient changes occur. This introduces potential for bias as a level of knowledge of the modelled domains could influence the results. It is however important to remember that the outputs from this process are maps, and therefore even though different individuals will generally produce

different results when splitting the ECDF plot by gradient, the same general trends will be obtained. Of each of the 3 methods, the ECDF methodology provides the greatest detail and opens the methodology to applications other than the identification of land contamination.

5.2 PTE Domains

Within this investigation, the maps produced using the ECDF technique were compared to the main factors known to control the distribution of elements across Northern Ireland in order to identify domains. The majority of domains were easily identified and the results correlated well with existing literature describing the distribution of elements in Northern Ireland (Young 2013 In Press).

Fig. 3 Domains identified for **a** arsenic **b** chromium, copper, nickel and vanadium and **c** lead based on the ECDF maps produced

Finalised domains are shown in Figure 3 for the elements under investigation: arsenic, chromium, copper, lead, nickel and vanadium. Similar controlling factors were identified for Cr, Cu, Ni and V, with elevated concentrations of these elements observed over areas of basalt bedrock geology creating a basalt domain. Reduced concentrations are seen in the Mourne Mountains Complex, associated with naturally occurring low concentrations of these elements in granites, creating the Mournes domain. Cr, Cu, Ni and V are known to be found at elevated concentrations in the Antrim Basalts (Barrat and Nesbitt 1996; Hill et al. 2001; Smith and McAlister 1995) and at reduced concentrations in granites (Wedepohl et al. 1978).

Concentrations of Cr, Cu, Ni and V were generally depleted in peat samples overlying all bedrock geologies except the basalts. Overlying the basalt formation, some areas of peat showed depleted Cr, Cu, Ni and V concentrations, while others (generally at lower topographical elevation) showed higher concentrations of each element in line with the basalt domain. This distribution is probably explained by the type of peatland and the land use activities taking place on it (Joint Nature Conservation Committee 2011). Lowland peats appear to have less of a control over element concentrations, meaning the basalts remain as the primary controlling factor and higher concentrations are observed. Upland peats, however, appear to exert a greater control with reduced concentrations being observed. It is also possible that the differing land use on the peat could be affecting its ability to function efficiently, however this subject would require further exploration. This distribution of Cr, Cu, Ni and V was also observed by Young (2013 In Press) and therefore a peat domain was selected which incorporated all areas of peat that do not overlie the basalts, along with areas of peat overlying the basalts that are associated with reduced concentrations of these elements. Depletion of Cr, Cu, Ni and V in this domain may reflect biogeochemical cycling of PTEs within the peats (Novak et al. 2011) but further research would be required to confirm this.

All the domains shown for lead are associated with elevated concentrations of the element. The Mourne Mountains Complex show elevated concentrations of lead (Mournes domain), fitting with known elevations of lead in granites (Krauskopf 1979). Elevated concentrations of lead were also associated with urban areas across Northern Ireland (urban domain). Pb is well known for its correlation with anthropogenic activity, and therefore urban centres (Albanese et al. 2011; Locutura and Bel-lan 2011). Identified sources of lead in urban environments include historical use of leaded fuel and lead in paint (Chirenje et al. 2004; Mielke and Zahran

2012; Mielke et al. 2011). A strong correlation was observed between areas of elevated topography with a covering of peat and elevated lead concentrations, as previously described by Young (2013 In Press). This is not surprising, as peat soils are well known for acting as historical records of atmospheric pollution, with higher heavy metal concentrations common in upper peat layers (Givelet et al. 2004; De Vleeschouwer et al. 2007). Chronologies of Pb deposition have been completed for a raised bog in Ireland (Coggins et al. 2006), showing elevated concentrations within the depth range (5-20cm) of the shallow Tellus samples. These topographically elevated areas of peat were separated out from the full peat dataset to form lead's peat domain.

Fig. 4 Mineral occurrences provided by GSNI shown for **a** lead on a map showing lead's mineralisation domain and **b** zinc, lead, gold and copper on a map showing arsenic's mineralisation domains (Crown Copyright)

Finally, a mineralisation domain was also identified for lead. This area was defined using the elevated concentrations of lead as determined on the ECDF map and the extent of the domain was corroborated by the strong correlation between lead mineral occurrences as provided by GSNI, shown in Figure 4a, areas of high to very high prospectivity potential (Lusty et al. 2012) and the mineralised domain identified for Pb. It is worth noting that lead mineral occurrences were identified by GSNI in the south east of Northern Ireland (Figure 4a), however as significantly elevated Pb concentrations were not recorded in that area a mineralised domain was not created in this region. In contrast, the area of lead mineralisation identified in the north west of Northern Ireland was in an area of elevated Pb concentrations. However, closer inspection of the spatial distribution of areas of elevated concentrations and comparison with peat maps showed that although mineralisation may be contributing to the elevated concentrations, their spatial distribution suggests that peat has a controlling role in accumulating this element, probably from atmospheric deposition.

The domains for arsenic were more difficult to define as there was greater uncertainty regarding the factors controlling the distribution of this element. The Shanmullagh formation, consisting of early Devonian age sandstone and mudstone (Mitchell 2004) contained slightly elevated arsenic concentrations, creating the Shanmullagh domain. Two other areas, thought to be associated with mineralisation, were also shown to contain elevated concentrations. These two mineralisation domains were defined in the same manner as the lead mineralisation domain, using the ECDF map, and were named mineralisation 1 and 2. As there are no shown arsenic occurrences on the mineral occurrences information provided by GSNI, a different approach was used to corroborate the extent of these areas. Arsenic is well known as a pathfinder for gold, and is therefore used in prospectivity investigations along with silver, gold, copper, lead, zinc, bismuth and barium (Lusty et al. 2009). A strong correlation is again shown between the mineralised zones identified in this study and mineral occurrences of Ag, Cu, Pb and Zn (Figure 4b), the only mineral occurrences from the previous list that were available from GSNI. The prospectivity for gold within all the identified As mineralised domains is again shown to be high (Lusty et al. 2009, 2012). An interesting, and possibly unexpected finding is the lack of an urban domain for arsenic. However, this was also the case in the calculation of NBCs for England (Ander et al. 2013a).

5.2.1 Domain Corroboration

Table 1 Experimental semivariogram modelled parameters for Ni, Cr, V, As, Cu and Pb concentrations as measured by XRF in shallow soils of Northern Ireland where C_0 = nugget effect, C_1 and C_2 = structured

component, a_1 and a_2 = ranges of influence, Total C_x = sill and $C_0/\text{Total } C_x$ = proportion of variance accounted for by C_0

The results of the geostatistical approach employed as corroboration are given in Table 1 for all the PTEs investigated. These show that the extent of Cr, Cu, Ni and V distributions in Northern Ireland are strongly controlled by the presence of basalts in the northeast of the region, with ranges (a) not exceeding the largest spatial extent of this geologic formation of approximately 90 kilometres (Figure 1b). Based on variography and source domain identification, elevated concentrations of these four trace elements in the region are attributable to geogenic sources on a spatial scale that exceeds other potential influences over the distribution of this element. However, approximately 13-35% of total variances in Cr, Cu, Ni and V spatial distributions are accounted for by the nugget variance (C_0) suggesting such proportions of variance may be accounted for by smaller scale processes not detected within the soil sampling resolution of the Tellus Survey (Table 1). Arsenic, by comparison, is controlled by a spatial function covering a smaller spatial extent than elements associated with the basalts, with a nugget effect accounting for approximately half of all variance in spatial distribution (48.1%). Lead exhibits the shortest range spatial function, characteristic of trace elements whose distributions are heavily influenced by small scale processes such as anthropogenic activity in urban areas. This trend is also supported by the large nugget effect for this element.

On the whole, these results confirm the main controlling factors identified for the PTEs investigated, especially the identification of a basalt domain for Cr, Cu, Ni and V. While the semi-variogram is very useful for identifying spatial controls over elevated element concentrations, reduced domain concentrations cannot be identified using this method.

5.3 Typical Threshold Values

5.3.1 Normal Background Concentrations

In order to assess what is a typical concentration of elements in Northern Irish soils, both the NBC statistical methodology developed by Cave et al. (2012), and the ULBL statistical methodology (Jarva et al. 2010) were applied to data within the defined domains.

Fig. 5 Outputs derived from NBC methodology using R-scripts developed by Cave et al. (2012) for nickel's basalt domain showing **a** histogram and **b** percentiles and relative uncertainty computed using the empirical, gaussian and robust methods

Figure 5 gives a visual example of how the NBC methodology was applied. The distribution of Ni within the basalt domain was assessed using a histogram; the values fell within the parameters set by Cave et al. (2012) (Figure 5a) allowing a Gaussian approach to be adopted for calculating percentiles (Figure 5b). A bootstrapping method was applied to calculate the uncertainty surrounding the percentile values (Figure 5b) and the 50th, 75th and 95th percentiles and their associated uncertainty are plotted in Figure 6.

Fig. 6 50th, 75th and 95th percentiles of each of the elements' domains along with their respective 95% upper and lower confidence limits (vertical lines shown), ULBL concentrations and SGV/GAC where R = residential, A =

allotment and C = commercial. Previous SGVs for Pb have been withdrawn; elsewhere SGVs/GACs that are not given are beyond the scale of the graphs

Elevation of Cr, Cu, Ni and V in the basalt domain is obvious from Figure 6, while reduced concentrations of these four elements are seen in the Mournes and peat domains. For Cr, the differences between the domains are maintained throughout the 50th, 75th and 95th percentiles. The upper 95% confidence limits of the 95th percentiles for the basalt, Mournes, peat and principal domains are 460, 84, 150 and 290 mg/kg respectively. The Mournes and peat domain contain substantially lower concentrations than those in the principal domain.

Similar results are shown for Cu at the 50th and 75th percentiles, but the 95th percentile shows a slight skew in Cu concentrations in the peat domain, with a higher concentration calculated at the 95th percentile in the peats than in the principal domain. The upper 95% confidence limits of the 95th percentile for the basalt, Mournes, peat and principal domains are 130, 41, 68 and 59 mg/kg. The distribution of Cu in the peat domain is heavily right-skewed, with a large presence of outliers. However, a log-transformation of the data brought it within the skewness limits set by Cave et al. (2012) and the gaussian approach was followed for calculating percentiles. Another possible explanation for this distribution is the existence of another controlling factor, other than solely peat substrate, which is contributing to the concentrations of Cu in this domain.

A large degree of uncertainty is associated with the values generated for Ni and V in the Mournes domain. For Ni, the 95th percentile was calculated as 24 mg/kg, with the lower and upper confidence limits calculated as 12 and 170 mg/kg respectively. The NBC value calculated for the Ni Mournes domain, of 170 mg/kg would therefore appear to be unrealistic, as the maximum value of Ni encountered in this domain was 37 mg/kg. The Mournes domain for Cu, Cr, Ni and V are based on the same 73 data points, making it the smallest domain, but still exceeding the 30 data points recommended in the NBC methodology (Cave et al. 2012). Reasonably large uncertainty is also calculated for vanadium, with the lower and upper confidence limits of the 95th percentile calculated as 39 and 174 mg/kg respectively. In comparison, much less uncertainty is shown for chromium and copper, where the differences between the upper and lower confidence limits for the 95th percentile are 40 and 22 mg/kg respectively. It seems that the distributions of these elements in the Mournes domain are responsible for the degree of uncertainty associated with the percentiles calculated. This may be due to the fact that the Mourne granite complex has several fabrics associated with fractionation of basaltic and crustal rock melts (Meighan et al. 1984; Stevenson and Bennett 2011). For V and Ni, a larger occurrence of outliers means that the Box-cox transformation was applied but higher uncertainties were still recorded for the 95th percentiles using this method. Fewer outliers present for Cr and Cu reduced the uncertainty associated with the percentiles, allowing for the calculation of more realistic NBCs.

The existence of outliers in vanadium's peat domain saw the application of a log-transformation in order to bring its distribution closer to normal. However, the outliers have still affected the calculation of the 95th percentile (120 mg/kg) and its confidence limits, while the 50th (27 mg/kg) and 75th (49mg/kg) percentiles appear to remain more representative. Nickel's peat domain contained even more outliers, meaning in this case the Box-cox transformation was required to bring the data within the necessary skewness limits (Cave et al. 2012). In this instance, the box-cox transformation appears to be more effective in reducing the effect of the

outliers, meaning the 50th, 75th and 95th percentiles (7mg/kg, 14mg/kg and 46 mg/kg) seem to give appropriate concentrations when compared to the mapped outputs (Figure 2).

For As, elevated concentrations are seen in the two mineralisation domains and the Shanmullagh domain when compared to the principal domain. If the 95th percentile is used as a comparison, the mineralisation 2 domain result is approximately five times greater than the 95th percentile for the principal domain. However, a large degree of uncertainty is associated with the 95th percentiles in the mineralisation 2 domain, as the lower and upper confidence limits range between 56 and 85 mg/kg respectively. The data for this domain were box-cox transformed and theoretical percentiles calculated based on the mean and standard deviation of the data set after an assessment of its distribution. However, the presence of one extreme outlier which lies over 60 mg/kg away from the remainder of the outliers is likely to be causing the extreme skew seen in the calculation of the 95th percentile. For the other three domains; mineralisation1, principal and Shanmullagh, the assessment of the data distribution led to robust percentiles and uncertainties being calculated. Although outliers are also present in these domains, none of them contain outliers as extreme as the one identified for the mineralisation 2 domain. This causes much smaller differences between the upper and lower 95th confidence limits than those identified for the mineralisation 2 domain.

For Pb, elevated concentrations are obvious in the urban domain, followed by the mineralisation, Mournes and peat domains. The lowest concentrations of Pb are found in the principal domain. Within the urban domain, reasonably large differences occur between the upper and lower confidence limits, as they range between 240 and 300 mg/kg for the 95th percentile. This is to be expected in the urban domain as anthropogenic sources of Pb increase the amount of outliers present in the data set, which in turn increases the uncertainty associated with the percentiles calculated. Although these values of lead are high compared to the principal domain for Northern Ireland, where the 95th percentile ranges between 71 and 77 mg/kg, they are not as high as those found in some other urban areas. Chirenje et al. (2004) reported a 95th percentile of Pb concentrations in Miami, USA, as 453 mg/kg.

Figure 6 shows both the strengths and the weaknesses of the NBC methodology. A large presence of outliers within the data set causes issues in the distribution assessment and ultimately in the uncertainty calculations. This stems from the overall distribution of the data, and the effectiveness of the transformation applied and is particularly obvious in the Mournes domain for Ni and V. It is important to note that the NBC methodology is not meant to be used for reduced concentration domains, which the Mournes domains for Ni and V are examples of. However, even for As and Pb where all the domains contain elevated concentrations, the high degree of uncertainty associated with some of the domains causes unrealistic concentrations at the upper 95% confidence limit of the 95th percentile, which is likely to pose difficulties if attempts are made to use NBCs during risk based assessment of contaminated sites rather than just identifying potentially contaminated land as legally defined in the UK as was originally intended. Also, this raises the question as to whether the UCL of the 95th percentile is an effective means of differentiating between diffuse and point contamination from anthropogenic sources.

From Figure 6, it is clear that the values shown for the 50th percentile provide a more realistic representation of the comparison between the domains, i.e. for Ni the 50th percentile shows elevated concentrations in the basalt

domain (100 mg/kg), widespread concentrations of 27 mg/kg in the principal domain and reduced concentrations of 4.0 and 7.2 mg/kg in the Mournes and peat domains respectively. However, this median value should not be taken as the typical threshold value as it doesn't fulfil the aim of TTVs as whilst median values are a central value for the domain, they don't allow for a differentiation between diffuse and point source anthropogenic contamination. However considering the 50th percentile may be effective within sectors other than the contaminated land sector, as the 50th percentile values can provide useful information on reduced concentration zones, where a possible depletion of essential elements, such as copper, could have consequences for industries such as agriculture.

5.3.2 Upper Limit of Geochemical Baseline Variation

TTVs calculated using the ULBL methodology are also shown on Figure 6. The NBC value (upper confidence interval on the 95th percentile) calculated for copper's peat domain (68 mg/kg) is higher than the value calculated for the principal domain (59 mg/kg) despite outputs from the ECDF domain identification method in Figure 2 suggesting that the peat is an area of depleted copper concentrations. In comparison, the ULBL method provides a more accurate representation of the values expected in reduced domains, with the peat, Mournes and principal domain containing ULBL concentrations of 47, 27 and 76 mg/kg respectively.

At elevated concentrations, both the ULBL and NBC methods appear to calculate similar values, with the ULBL method generally calculating slightly lower values for Pb. With regard to the elevated concentrations in the basalt domain for Cr, Cu, Ni and V, the ULBL method calculates concentrations that are at least 20% higher than the respective NBC. This is probably due to the fact that the distribution of these elements in the basalt domain is relatively homogeneous and therefore closer to a normal distribution. When the boxplot is used to identify outlying values for a more normal distribution, fewer values will be identified and so a higher typical threshold value will be set using this method. Although set methods are used in the NBC methodology depending on the distribution of the data, a major strength of the boxplot is it's resistance to different types of distribution.

5.3.3 Comparison with relevant criteria

Figure 6 also provides details of SGVs and GACs where they are available for the elements. SGVs and GACs are used to "represent cautious estimates of levels of contaminants in soil at which there is considered to be no risk to health or, at most, a minimal risk to health" (Defra 2012). Therefore they are based on a different approach than that behind the NBC methodology where the aim is to identify sites where "if nothing is done, there is a significant possibility of significant harm" (Ander et al. 2013a). However, a comparison can still be drawn between the values. Figure 6 highlights certain domains for a number of the elements where the typical threshold values are higher than the reference values for residential, and in some cases allotment end uses. Therefore, depending on the size of the exceedance, surpassing these SGV values could suggest a possible risk to human health. For As, the residential SGV of 32 mg/kg (Martin et al. 2009a) is narrowly exceeded in the mineralisation 2 domain and is therefore unlikely to pose significant risks to human health. GAC for Cr-VI are shown on Figure 6 (Nathanail et al. 2009), with exceedance of these concentrations shown in all domains. However, no distinction was drawn between Cr-III and Cr-VI in the Tellus survey. For Ni, the basalt domain shows a significant exceedance of the residential SGV (130 mg/kg) (Martin et al. 2009b) using both the NBC (200 mg/kg) and the ULBL (250 mg/kg) methods, however recent studies by Barsby et al. (2012), Cox et al.

(2013) and Palmer et al. (2013) indicate the oral bioaccessibility of Ni in these soils is relatively low (1% to 44%). The NBC for Ni in the Mournes domain, whilst high (170 mg/kg), is not a representative value when compared to mapped outputs (Figure 2) and all values for Cu fall within the GAC (Nathanail et al. 2009) making them unlikely to pose significant risks to human health. The NBC for vanadium in the Mournes domain again is unrepresentatively high, with the ULBL method calculating a value of 46 mg/kg which appears to be more representative of mapped outputs. All vanadium results exceed the allotment GAC with values for the basalt, peat and principal domains also exceeding the residential GAC (Nathanail et al. 2009). However bioaccessibility testing reported in Barsby et al. (2012) and Palmer et al. (2013) suggests that only a small fraction of total V in these areas is bioaccessible (8%).

6 Conclusions

In terms of domain identification, the three methods exhibit specific advantages and disadvantages. The k-means technique provided useful results in the determination of elevated domains but its applicability in practice would be limited as it cannot be used to define reduced concentration domains. The boxplot and ECDF methods both allowed identification of elevated and reduced concentration domains. However, the boxplot method splits the distribution at arbitrary values, whereas changes in gradient linked to different data distributions within the overall data set are used to divide the ECDF graph. Splitting the ECDF graph requires a level of interpretation by the individual completing the work, which introduces potential for bias as a level of knowledge of the modelled domains could influence the results. Of the 3 methods, the ECDF methodology provides the greatest amount of detail and opens the methodology to other practical applications rather than just identification of land contamination. However, choice of method may ultimately lie with the decision maker as whatever method they choose may depend on the original goals behind the use of this methodology.

The NBC methodology has been developed to sit within a specific legislative framework. By defining the NBC as the upper 95% confidence limit of the 95th percentile it generates a question as to how conservative the approach taken is. In contrast to this, the ULBL methodology generates the maximum number of outliers by using a boxplot of non-transformed data, in order to generate the lowest ULBL concentration. This is demonstrated in Figure 6, where generally the ULBL values calculated are slightly lower than the NBC concentrations. A notable exception to this is for Cr, Cu, Ni and V in the basalts domain, where on all four occasions the ULBL method calculated higher concentrations than the NBC methodology. The largest difference was for Cu, where the ULBL method calculated a concentration 28% higher than the NBC. As well as this, the distribution of the data within each of the domains seems to have a large control over the amount of uncertainty calculated for the relevant percentiles in the NBC method, particularly where a large amount of outliers are present. The transformations applied mainly account for the skewed distributions, but examples remain where the data transformation does not seem to be fully effective (vanadium's peat domain). It is clear from the previous discussion that both methods have their strengths, however in general the ULBL concentrations provide more realistic concentrations for typical threshold values as defined in this study, across the area and elements shown.

An interesting investigation, following on from this work, would be to consider the geographic location of each of the outliers identified using the ULBL method. If specific sources of elements could be identified as causing

the elevated concentrations of the outlying values, then clarification of how effectively the method discriminates between anthropogenic diffuse and point source concentrations of elements could be gained.

One of the primary aims of this research was to investigate soil geochemical data for Northern Ireland, and determine an output in the form of relevant TTVs which define the boundary between geogenic and diffuse anthropogenic source contributions to soil, and those associated with point sources. In this respect, the following is suggested for use;

- ECDF mapping method for identifying the main controls over PTE concentration distributions and allowing the identification of domains,
- Calculation of TTVs using the ULBL method (currently employed in Finland) within each of the defined domains.

These values will be of interest to a number of parties, as they indicate what a “typical” concentration of an element would be within a defined geographical area. These values should be considered alongside the risk that each of the PTEs pose in these areas, in order to determine potential risk to receptors.

7 Acknowledgements

Alex Donald of the Geological Survey of Northern Ireland (GSNI) is thanked for arranging access to the Tellus data. Many thanks to GSNI for providing their superficial and bedrock geology maps (Crown Copyright), and their information on mineral occurrences (Crown Copyright). The Tellus Project was funded by the Department of Enterprise Trade and Investment and by the Rural Development Programme through the Northern Ireland Programme for Building Sustainable Prosperity. This research is supported by the EU INTERREG IVA-funded Tellus Border project. The views and opinions expressed in this research report do not necessarily reflect those of the European Commission or the SEUPB. The authors declare that they have no conflict of interest.

8 References

- Ajmone-Marsan, F. et al. 2008. “Metals in Particle-Size Fractions of the Soils of Five European Cities.” *Environmental Pollution* 152:73–81. Retrieved December 19, 2012 (<http://www.ncbi.nlm.nih.gov/pubmed/17602808>).
- Albanese, Stefano et al. 2011. “Advancements in Urban Geochemical Mapping of the Naples Metropolitan Area: Colour Composite Maps and Results from an Urban Brownfield Site.” Pp. 410–24 in *Mapping the Chemical Environment of Urban Areas*, edited by Christopher C Johnson, Alecos Demetriades, Juan Locutura, and R T Ottesen. Oxford: Wiley-Blackwell.
- Ander, E. Louise, Christopher C. Johnson, et al. 2013a. “Methodology for the Determination of Normal Background Concentrations of Contaminants in English Soil.” *The Science of the total environment* 454-455:604–18. Retrieved May 7, 2013 (<http://www.ncbi.nlm.nih.gov/pubmed/23583985>).
- Ander, E. Louise, Mark R. Cave, and Christopher C. Johnson. 2013b. “Normal Background Concentrations of Contaminants in the Soils of Wales. Exploratory Data Analysis and Statistical Methods.” *British Geological Survey Commissioned Report CR/12/107*:144pp.

609 Ander, EL, Mark R. Cave, Christopher C. Johnson, and B. Palumbo-Roe. 2011. "Normal Background
610 Concentrations of Contaminants in the Soils of England. Available Data and Data Exploration." *British
611 Geological Survey Commissioned Report CR/11/145:124pp.*

612 APAT-ISS. 2006. "Protocollo Operativo per La Determinazione Dei Valori Di Fondo Di Metalli/metalloidi Nei
613 Suoli Dei Siti D'interesse Nazionale." *Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici and
614 Istituto Superiore di Sanita* Revisione 0.

615 Barrat, JA, and RW Nesbitt. 1996. "Geochemistry of the Tertiary Volcanism of Northern Ireland." *Chemical
616 Geology* 129:15–38.

617 Barsby, Amy et al. 2012. "Bioaccessibility of Trace Elements in Soils in Northern Ireland." *The Science of the
618 total environment* 433:398–417. Retrieved November 29, 2012
619 (<http://www.ncbi.nlm.nih.gov/pubmed/22819891>).

620 British Standards. 2011. *Soil Quality — Guidance on the Determination of Background Values BS EN ISO
621 19258:2011.*

622 Cave, Mark R., Christopher C. Johnson, EL Ander, and B. Palumbo-Roe. 2012. "Methodology for the
623 Determination of Normal Background Contaminant Concentrations in English Soils." *British Geological
624 Survey Commissioned Report CR/12/003:42pp.*

625 Chirenje, T. et al. 2003. "Arsenic Distribution in Florida Urban Soils: Comparison between Gainesville and
626 Miami." *Journal of Environmental Quality* 32:109–19. Retrieved
627 (<http://www.ncbi.nlm.nih.gov/pubmed/12549549>).

628 Chirenje, Tait, L. Q. Ma, M. Reeves, and M. Szulczewski. 2004. "Lead Distribution in near-Surface Soils of
629 Two Florida Cities: Gainesville and Miami." *Geoderma* 119:113–20. Retrieved February 8, 2012
630 (<http://linkinghub.elsevier.com/retrieve/pii/S0016706103002441>).

631 Coggins, A. M., S. G. Jennings, and R. Ebinghaus. 2006. "Accumulation Rates of the Heavy Metals Lead,
632 Mercury and Cadmium in Ombrotrophic Peatlands in the West of Ireland." *Atmospheric Environment*
633 40:260–78. Retrieved October 25, 2013
634 (<http://linkinghub.elsevier.com/retrieve/pii/S135223100500899X>).

635 Communities and Local Government. 2007. "Generalised Land Use Database Statistics for England 2005."
636 Product Code 06CSRG04342.

637 Cox, Siobhan F. et al. 2013. "The Importance of Solid-Phase Distribution on the Oral Bioaccessibility of Ni and
638 Cr in Soils Overlying Palaeogene Basalt Lavas, Northern Ireland." *Environmental Geochemistry and
639 Health* 35:553–67. Retrieved August 24, 2013 (<http://www.ncbi.nlm.nih.gov/pubmed/23821222>).

640 Cruickshank, JG. 1997. *Soil and environment:Northern Ireland.* edited by JG Cruickshank.

641 Davies, Helen, and Sally Walker. 2013. *Strategic Planning Policy Statement (SPPS) for Northern Ireland:
642 Strategic Environmental Assessment (SEA) Scoping Report.* Leeds.

643 Department for Environment Food and Rural Affairs (Defra). 2012. *Environmental Protection Act 1990 : Part
644 2A Contaminated Land Statutory Guidance.* Retrieved
645 (<http://www.defra.gov.uk/environment/quality/land/>).

646 Deutsch, Clayton V, and Andre G. Journel. 1998. *GSLIB: Geostatistical Software Library and User's Guide.*
647 Second. New York: Oxford University Press.

648 Díez, M., M. Simón, C. Dorronsoro, I. García, and F. Martín. 2007. "Background Arsenic Concentrations in
649 Southeastern Spanish Soils." *Science of the Total Environment* 378:5–12. Retrieved
650 (<http://edafologia.ugr.es/comun/trabajos/total07diez.pdf>).

651 ESRI (Environmental Systems Resource Institute). 2009. "ArcMap 10."

652 European Environment Agency. 2012. "Corine Land Cover 2006 Seamless Vector Data." Version 16 (04/2012).
 653 Retrieved (<http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version-2#tab-additional->
 654 [information](http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version-2#tab-additional-)).

655 Givélet, Nicolas et al. 2004. "Suggested Protocol for Collecting, Handling and Preparing Peat Cores and Peat
 656 Samples for Physical, Chemical, Mineralogical and Isotopic Analyses." *Journal of environmental*
 657 *monitoring : JEM* 6:481–92. Retrieved (<http://www.ncbi.nlm.nih.gov/pubmed/15152318>).

658 Goodale, Christine L., John D. Aber, and Scott V Ollinger. 1998. "Mapping Monthly Precipitation,
 659 Temperature, and Solar Radiation for Ireland with Polynomial Regression and a Digital Elevation Model."
 660 *Climate Research* 10:35–49.

661 Green, KA, S. Caven, and T. R. Lister. 2010. *Tellus Soil Geochemistry - Quality Assessment and Map*
 662 *Production of ICP Data*.

663 Hartigan, J. A., and A. Wong. 1979. "A K-Means Clustering Algorithm." *Journal of the Royal Statistical*
 664 *Society. Series C (Applied Statistics)* 28(1):100–108.

665 Hawkes, HE, and JS Webb. 1962. *Geochemistry in Mineral Exploration*. New York: Harper & Row Publishers.

666 Hill, IG, RH Worden, and IG Meighan. 2001. "Formation of Interbasaltic Laterite Horizons in NE Ireland by
 667 Early Tertiary Weathering Processes." Pp. 339–48 in *Geologists' Association*, vol. 112. The Geologists'
 668 Association. Retrieved September 25, 2013
 669 (<http://linkinghub.elsevier.com/retrieve/pii/S0016787801800134>).

670 Jain, A. K., M. N. Murty, and P. J. Flynn. 1999. "Data Clustering: A Review." *ACM Computing Surveys*
 671 31(3):264–323.

672 Jarva, Jaana, Timo Tarvainen, Jussi Reinikainen, and Mikael Eklund. 2010. "TAPIR - Finnish National
 673 Geochemical Baseline Database." *Science of the Total Environment* 408:4385–95.

674 Johnson, Christopher C., and E. Louise Ander. 2008. "Urban Geochemical Mapping Studies: How and Why We
 675 Do Them." *Environmental geochemistry and health* 30:511–30. Retrieved February 11, 2014
 676 (<http://www.ncbi.nlm.nih.gov/pubmed/18563589>).

677 Joint Nature Conservation Committee. 2011. *Towards an Assessment of the State of UK Peatlands*.

678 Jordan, C. 2001. *The Soil Geochemical Atlas of Northern Ireland*.

679 Journel, AG, and Ch J. Huijbregts. 1978. *Mining Geostatistics*. Orlando: Academic Press.

680 Kelepertsis, A., A. Argyraki, and D. Alexakis. 2006. "Multivariate Statistics and Spatial Interpretation of
 681 Geochemical Data for Assessing Soil Contamination by Potentially Toxic Elements in the Mining Area of
 682 Stratoni, North Greece." *Geochemistry: Exploration, Environment, Analysis* 6:349–55. Retrieved
 683 (<http://geea.geoscienceworld.org/cgi/doi/10.1144/1467-7873/05-101>).

684 Krauskopf, Konrad B. 1979. *Introduction to Geochemistry*. Second. edited by Donald C Jackson. McGraw-Hill
 685 Book Company.

686 Lloyd, CD. 2007. *Local Models for Spatial Analysis*. London: CRC Press: Taylor and Francis Group.

687 Locutura, Juan, and Alejandro Bel-lan. 2011. "Systematic Urban Geochemistry of Madrid, Spain, Based on
 688 Soils and Dust." Pp. 307–47 in *Mapping the Chemical Environment of Urban Areas*, edited by

- 689 Christopher C Johnson, Alecos Demetriades, Juan Locutura, and Rolf Tore Ottesen. Oxford: Wiley-
690 Blackwell.
- 691 Lusty, PAJ, PM McDonnell, AG Gunn, BC Chacksfield, and M. Cooper. 2009. "Gold Potential of the Dalradian
692 Rocks of North-West Northern Ireland: Prospectivity Analysis Using Tellus Data." *British Geological*
693 *Survey Internal Report* OR/08/39:74pp.
- 694 Lusty, PAJ, C. Scheib, AG Gunn, and ASD Walker. 2012. "Reconnaissance-Scale Prospectivity Analysis for
695 Gold Mineralisation in the Southern Uplands-Down-Longford Terrane , Northern Ireland." *Natural*
696 *Resources Research* 21(3):359–82.
- 697 Martin, I., R. De Burca, and H. Morgan. 2009a. *Soil Guideline Values for Inorganic Arsenic in Soil*. Retrieved
698 (<http://www.environment-agency.gov.uk/static/documents/Research/SCHO0409BPVY-e-e.pdf>).
- 699 Martin, I., H. Morgan, C. Jones, E. Waterfall, and J. Jeffries. 2009b. *Soil Guideline Values for Nickel in Soil*.
700 Retrieved ([http://www.environment-agency.gov.uk/static/documents/Research/SCHO0409BPWB-e-](http://www.environment-agency.gov.uk/static/documents/Research/SCHO0409BPWB-e-e.pdf)
701 [e.pdf](http://www.environment-agency.gov.uk/static/documents/Research/SCHO0409BPWB-e-e.pdf)).
- 702 Matheron, G. 1965. *The Theory of Regionalised Variables and Their Estimation*. Paris: Masson.
- 703 Matschullat, J., R. Ottenstein, and C. Reimann. 2000. "Geochemical Background – Can We Calculate It?"
704 *Environmental Geology* 39(9):990–1000.
- 705 McKinley, JM, CD Lloyd, and AH Ruffell. 2004. "Use of Variography in Permeability Characterization of
706 Visually Homogeneous Sandstone Reservoirs with Examples from Outcrop Studies." *Mathematical*
707 *Geology* 36(7):761–79.
- 708 Meighan, I. G., D. Gibson, and D. N. Hood. 1984. "Some Aspects of Tertiary Acid Magmatism in NE Ireland."
709 *Mineralogical Magazine* 48:351–63.
- 710 Mielke, Howard W., Jan Alexander, Marianne Langedal, and Rolf Tore Ottesen. 2011. "Children, Soils and
711 Health: How Do Polluted Soils Influence Children's Health?" Pp. 134–50 in *Mapping the Chemical*
712 *Environment of Urban Areas*, edited by Christopher C Johnson, Alecos Demetriades, Juan Locutura, and
713 R T Ottesen. Oxford: Wiley-Blackwell.
- 714 Mielke, Howard W., and Sammy Zahran. 2012. "The Urban Rise and Fall of Air Lead (Pb) and the Latent Surge
715 and Retreat of Societal Violence." *Environment international* 43:48–55. Retrieved April 17, 2012
716 (<http://www.ncbi.nlm.nih.gov/pubmed/22484219>).
- 717 Ministry of the Environment Finland. 2007. *Government Decree on the Assessment of Soil Contamination and*
718 *Remediation Needs (214/2007)*.
- 719 Mitchell, WI, ed. 2004. *The Geology of Northern Ireland*. Second Edi. Belfast: Geological Survey of Northern
720 Ireland.
- 721 Nathanail, P. et al. 2009. *The LQM/CIEH Generic Assessment Criteria for Human Health Risk Assessment*. 2nd
722 Editio.
- 723 Northern Ireland Statistics & Research Agency. 2013. *Population and Migration Estimates Northern Ireland*
724 *(2012) – Statistical Report*.
- 725 Novak, Martin et al. 2011. "Experimental Evidence for Mobility/immobility of Metals in Peat." *Environmental*
726 *science & technology* 45:7180–87. Retrieved (<http://www.ncbi.nlm.nih.gov/pubmed/21761934>).
- 727 Palmer, Sherry, Ulrich Ofterdinger, Jennifer M. McKinley, Siobhan Cox, and Amy Barsby. 2013. "Correlation
728 Analysis as a Tool to Investigate the Bioaccessibility of Nickel, Vanadium and Zinc in Northern Ireland

729 Soils." *Environmental geochemistry and health* 35:569–84. Retrieved September 6, 2013
730 (<http://www.ncbi.nlm.nih.gov/pubmed/23793447>).

731 Parnell, J. et al. 2000. "Regional Fluid Flow and Gold Mineralisation in the Dalradian of the Sperrin Mountains,
732 Northern Ireland." *Economic Geology* 95(7):1389–1416.

733 Paterson, E., W. Towers, J. R. Bacon, and M. Jones. 2003. *Background Levels of Contaminants in Scottish*
734 *Soils*.

735 R Core Team. 2013. "R: A Language and Environment for Statistical Computing." Retrieved ([http://www.r-](http://www.r-project.org)
736 [project.org](http://www.r-project.org)).

737 Ramos-Miras, JJ, L. Roca-Perez, M. Guzmán-Palomino, R. Boluda, and C. Gil. 2011. "Background Levels and
738 Baseline Values of Available Heavy Metals in Mediterranean Greenhouse Soils (Spain)." *Journal of*
739 *Geochemical Exploration* 110:186–92. Retrieved March 26, 2013
740 (<http://linkinghub.elsevier.com/retrieve/pii/S0375674211000884>).

741 Reimann, Clemens. 2005. "Geochemical Mapping - Technique or Art." *Geochemistry: Exploration,*
742 *Environment, Analysis* 5(4):359–70.

743 Reimann, Clemens, Peter Filzmoser, and Robert G. Garrett. 2005. "Background and Threshold: Critical
744 Comparison of Methods of Determination." *The Science of the total environment* 346:1–16. Retrieved
745 March 12, 2012 (<http://www.ncbi.nlm.nih.gov/pubmed/15993678>).

746 Reimann, Clemens, Peter Filzmoser, Robert G. Garrett, and Rudolf Dutter. 2008. *Statistical Data Analysis*
747 *Explained: Applied Environmental Statistics with R*. Chichester: John Wiley & Sons Ltd.

748 Reimann, Clemens, and Robert G. Garrett. 2005. "Geochemical Background--Concept and Reality." *The*
749 *Science of the total environment* 350:12–27. Retrieved October 19, 2012
750 (<http://www.ncbi.nlm.nih.gov/pubmed/15890388>).

751 Rodrigues, S. et al. 2006. "Mercury in Urban Soils: A Comparison of Local Spatial Variability in Six European
752 Cities." *The Science of the total environment* 368:926–36. Retrieved November 21, 2012
753 (<http://www.ncbi.nlm.nih.gov/pubmed/16750244>).

754 Rodrigues, S. M., M. E. Pereira, E. Ferreira da Silva, A. S. Hursthouse, and A. C. Duarte. 2009. "A Review of
755 Regulatory Decisions for Environmental Protection: Part I - Challenges in the Implementation of National
756 Soil Policies." *Environment international* 35:202–13. Retrieved March 19, 2013
757 (<http://www.ncbi.nlm.nih.gov/pubmed/18817974>).

758 Romesburg, H. Charles. 2004. *Cluster Analysis for Researchers*. North Carolina: Lulu Press.

759 Salminen, R., and T. Tarvainen. 1997. "The Problem of Defining Geochemical Baselines. A Case Study of
760 Selected Elements and Geological Materials in Finland." *Journal of Geochemical Exploration* 60:91–98.
761 Retrieved (<http://linkinghub.elsevier.com/retrieve/pii/S0375674297000289>).

762 Shotyk, William. 1996. "Peat Bog Archives of Atmospheric Metal Deposition: Geochemical Evaluation of Peat
763 Profiles, Natural Variations in Metal Concentrations, and Metal Enrichment Factors." *Environmental*
764 *Reviews* 4(2):149–83.

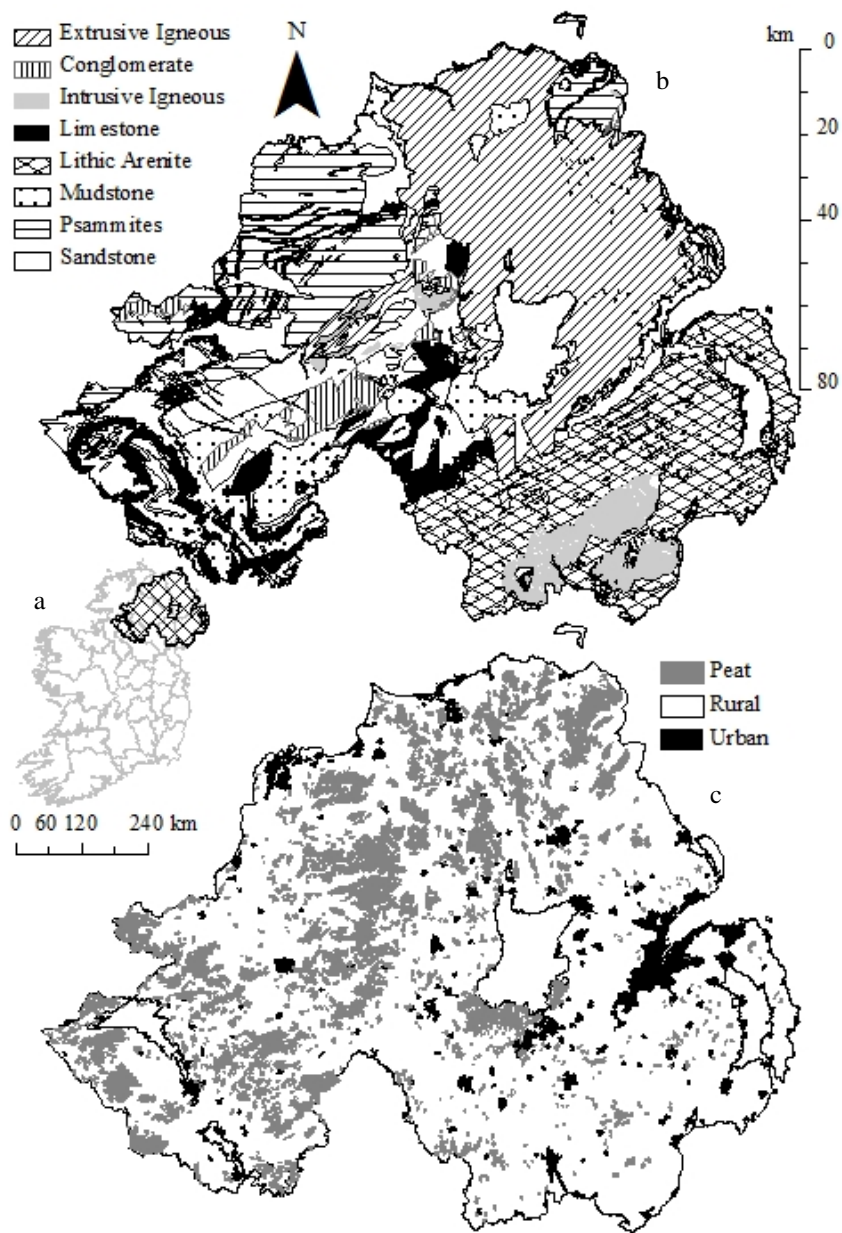
765 Sinclair, AJ. 1974. "Selection of Threshold Values in Geochemical Data Using Probability Graphs." *Journal of*
766 *Geochemical Exploration* 3:129–49.

767 Smith, BJ, and JJ McAlister. 1995. "Mineralogy , Chemistry and Palaeoenvironmental Significance of an Early
768 Tertiary Terra Rossa from Northern Ireland : A Preliminary Review." *Geomorphology* 12:63–73.

- Smyth, Dermot. 2007. "Methods Used in the Tellus Geochemical Mapping of Northern Ireland." *British Geological Survey Open Report* OR/07/022:90pp.
- Stevenson, C. T. E., and N. Bennett. 2011. "The Emplacement of the Palaeogene Mourne Granite Centres, Northern Ireland: New Results from the Western Mourne Centre." *Journal of the Geological Society* 168:831–36. Retrieved September 30, 2013 (<http://jgs.lyellcollection.org/cgi/doi/10.1144/0016-76492010-123>).
- Templ, Matthias, Peter Filzmoser, and Clemens Reimann. 2008. "Cluster Analysis Applied to Regional Geochemical Data: Problems and Possibilities." *Applied Geochemistry* 23(8):2198–2213. Retrieved October 29, 2012 (<http://linkinghub.elsevier.com/retrieve/pii/S088329270800125X>).
- De Vleeschouwer, François et al. 2007. "Atmospheric Lead and Heavy Metal Pollution Records from a Belgian Peat Bog Spanning the Last Two Millennia: Human Impact on a Regional to Global Scale." *The Science of the total environment* 377:282–95. Retrieved January 8, 2014 (<http://www.ncbi.nlm.nih.gov/pubmed/17379271>).
- Wedepohl, K. H., C. W. Correns, D. M. Shaw, K. K. Turekian, and J. Zemmann, eds. 1978. *Handbook of Geochemistry*. II–4th ed. Springer-Verlag Berlin.
- Wong, Coby S. C., Xiangdong Li, and Iain Thornton. 2006. "Urban Environmental Geochemistry of Trace Metals." *Environmental Pollution* 142:1–16. Retrieved February 7, 2013 (<http://www.ncbi.nlm.nih.gov/pubmed/16297517>).
- Young, ME. 2013. *A Guide to the Tellus Data*. In Press. edited by AW Donald.
- Zhang, Chaosheng, Crawford Jordan, and Alex Higgins. 2007. "Using Neighbourhood Statistics and GIS to Quantify and Visualize Spatial Variation in Geochemical Variables: An Example Using Ni Concentrations in the Topsoils of Northern Ireland." *Geoderma* 137:466–76. Retrieved October 19, 2012 (<http://linkinghub.elsevier.com/retrieve/pii/S0016706106003053>).

794 **Figures**

795 **Fig. 1**

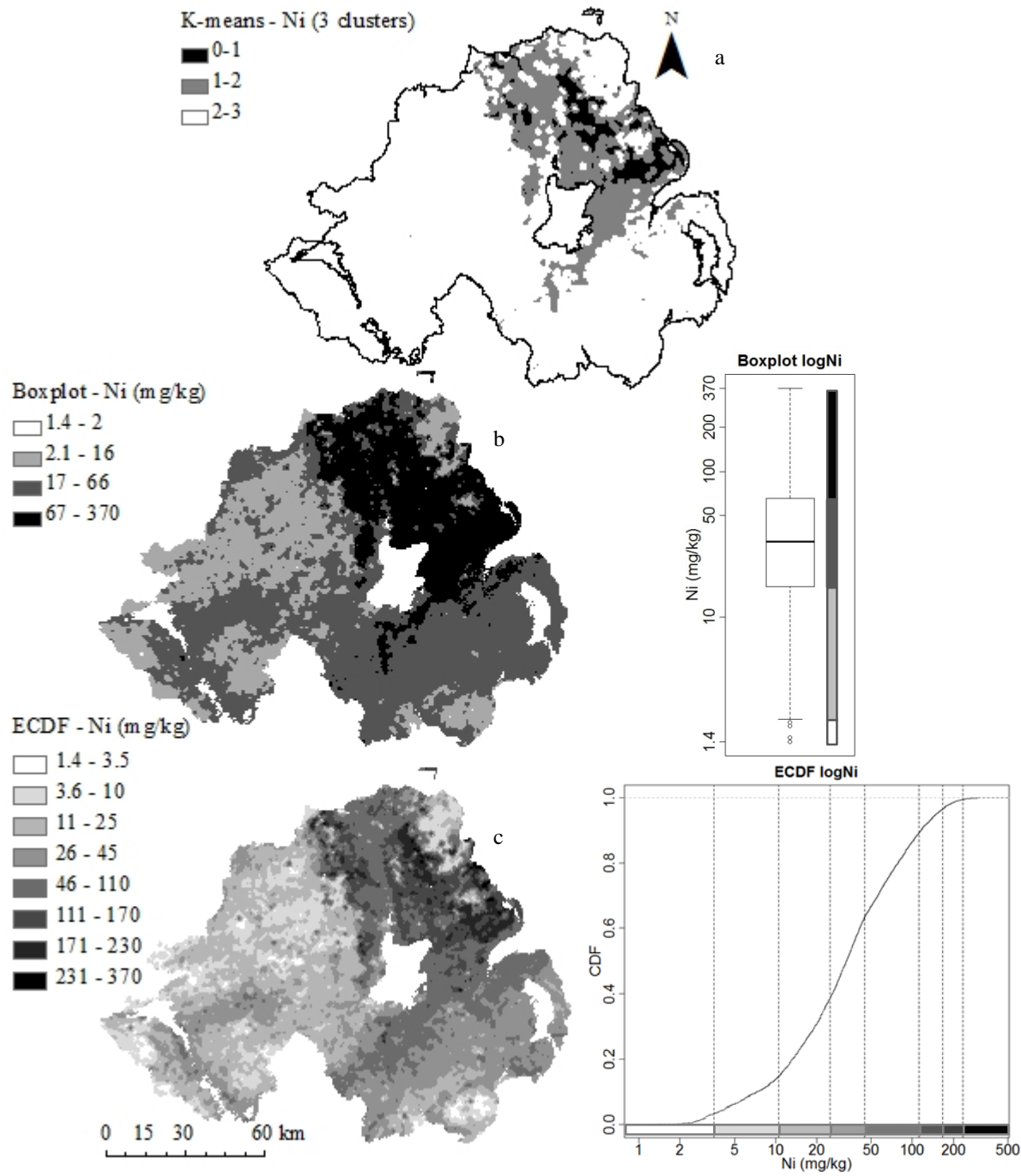


796

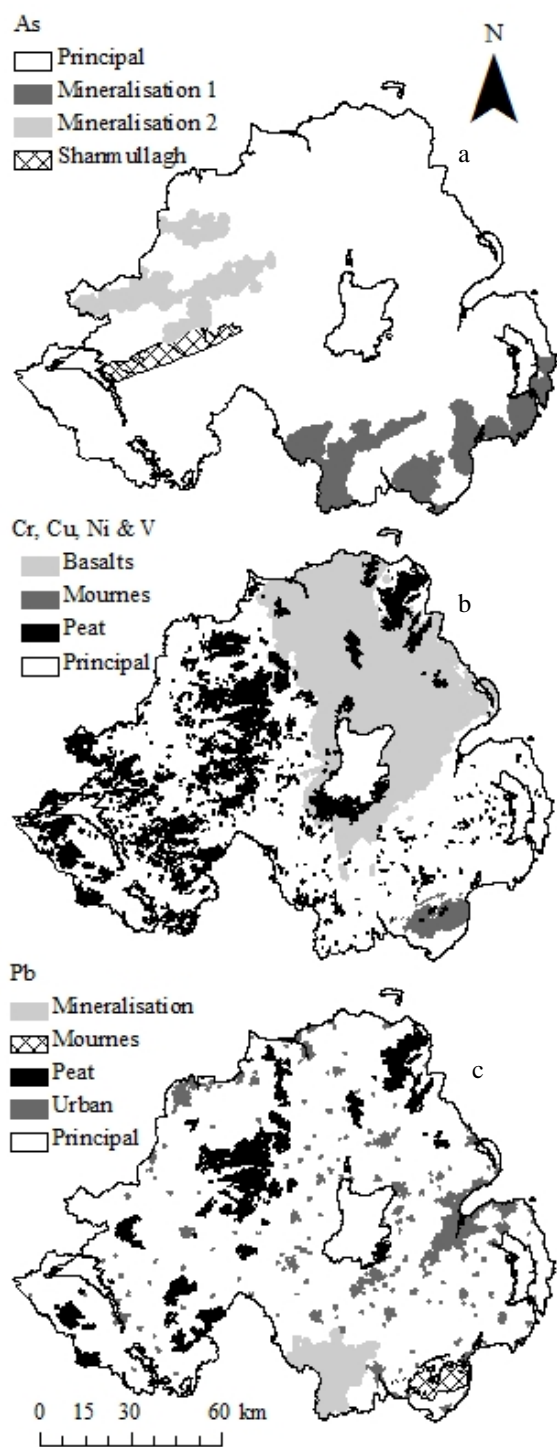
797

798 Fig. 2

799



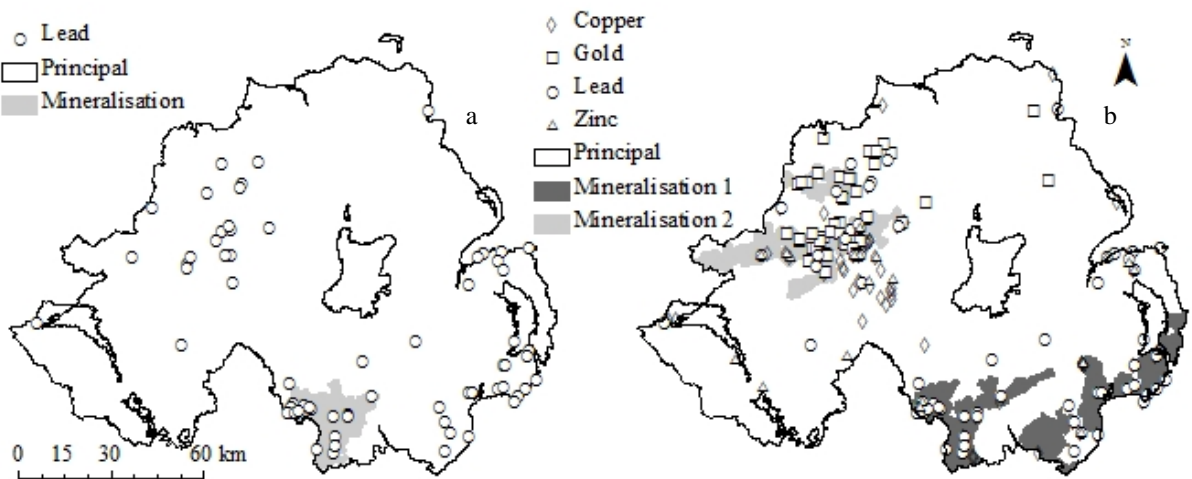
800 Fig. 3



801

802

803 **Fig. 4**

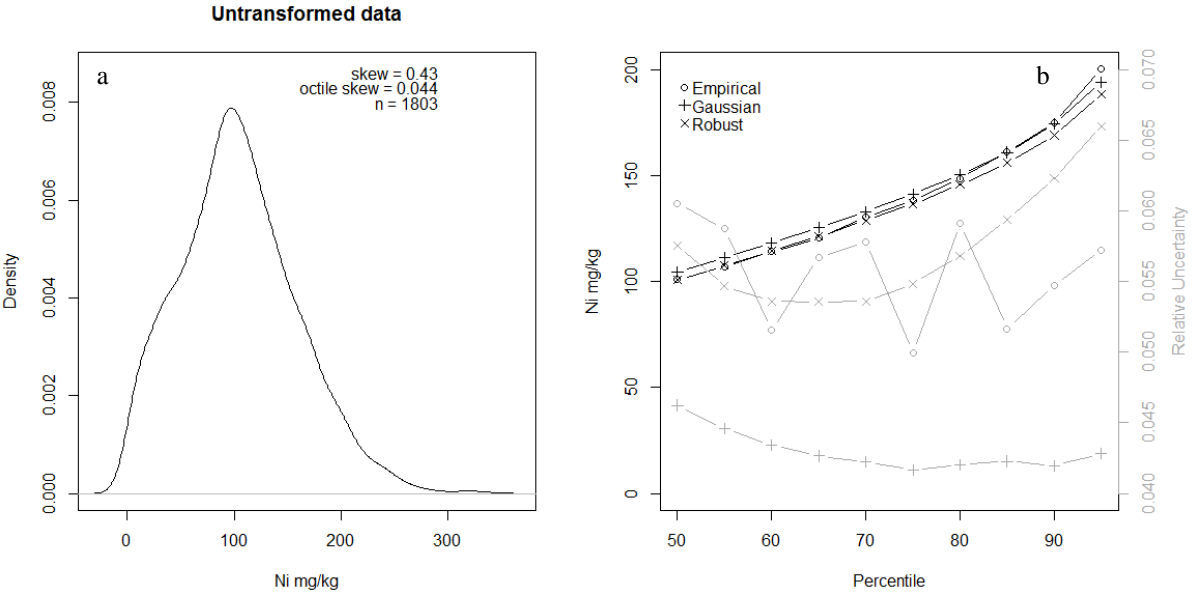


806 **Table 1**

<i>Element</i>	C_0	C_1	C_2	a_1 (m)	a_2 (m)	Total C_x	C_0 /Total C_x (%)
Ni	696.6	207.0	1,948.4	90,093.8	73,881.9	2,852.0	24.4
Cr	2,223.2	11,604.0	3,063.8	86,790.1	7,038.5	16,891.0	13.2
V	697.8	3,014.0	1,257.2	74,193.7	13,816.3	4,969.0	14.0
As	53.5	57.8	--	6,569.9	--	111.3	48.1
Cu	448.4	366.5	462.9	10,882.7	77,709.2	1,277.8	35.1
Pb [†]	55,056.6	12,324.1	--	17,736.0	--	67,380.7	81.7

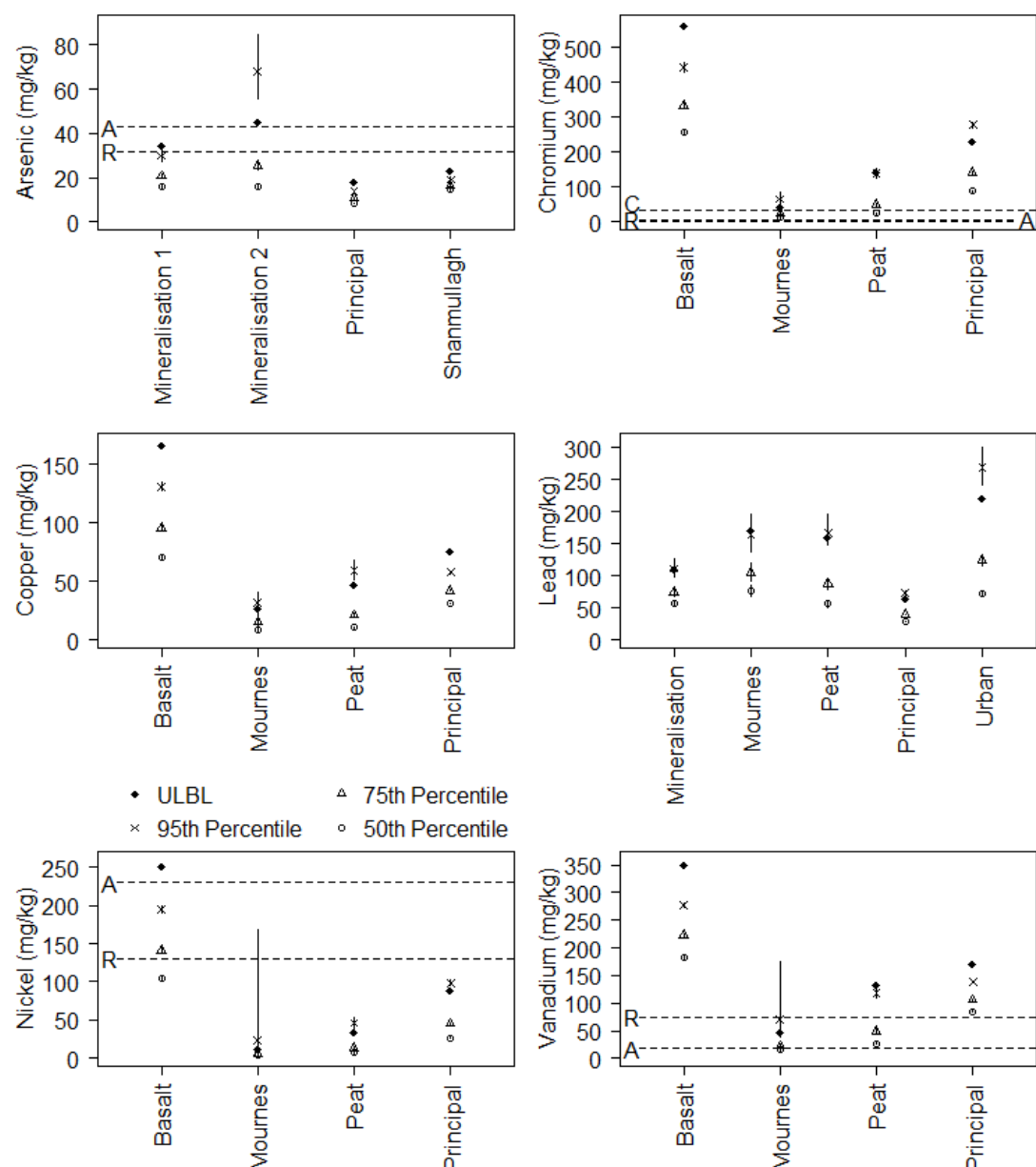
807 [†](Barsby et al. 2012)

808 **Fig. 5**



809

810



812

813

Supporting Information

Table S1 50th, 75th, 95th percentiles and ULBLs for As, Cr, Cu, Pb, Ni and V where n = number of data points in each domain, LCL = 95% lower confidence limit and UCL = 95% upper confidence limit

Arsenic (mg/kg)	n	50th	50th LCL	50th UCL	75th	75th LCL	75th UCL	95th	95th LCL	95th UCL	ULBL
Mineralisation 1	528	16	16	17	21	20	22	30	28	32	34
Mineralisation 2	351	16	15	17	26	23	28	68	56	85	45
Principal	7015	8.4	8.3	8.5	11	10	11	14	14	14	18
Shanmullagh	133	15	14	15	16	15	17	19	18	21	23

Chromium (mg/kg)	n	50th	50th LCL	50th UCL	75th	75th LCL	75th UCL	95th	95th LCL	95th UCL	ULBL
Basalt	1803	260	250	270	330	320	340	440	430	460	560
Mournes	73	13	11	17	23	19	37	65	43	84	43
Peat	713	24	22	26	50	45	54	140	120	150	140
Principal	5439	89	87	91	140	140	140	280	270	290	230

Copper (mg/kg)	n	50th	50th LCL	50th UCL	75th	75th LCL	75th UCL	95th	95th LCL	95th UCL	ULBL
Basalt	1803	71	69	73	96	93	98	130	130	130	170
Mournes	73	9.3	8.1	11	15	12	18	32	18	41	27
Peat	713	11	9.9	12	22	20	24	59	52	68	47
Principal	5439	32	31	32	42	42	43	58	57	59	76

Lead (mg/kg)	n	50th	50th LCL	50th UCL	75th	75th LCL	75th UCL	95th	95th LCL	95th UCL	ULBL
Mineralisation	208	57	53	60	75	69	81	110	99	130	110
Mournes	75	76	69	85	100	92	120	170	140	200	170
Peat	370	56	50	60	88	80	97	170	150	200	160
Principal	6394	29	28	29	40	39	41	74	71	77	63
Urban	981	73	70	77	130	120	130	270	240	300	220

Nickel (mg/kg)	n	50th	50th LCL	50th UCL	75th	75th LCL	75th UCL	95th	95th LCL	95th UCL	ULBL
Basalt	1803	100	100	110	140	140	140	190	190	200	250
Mournes	73	4.0	3.6	4.5	6.0	5.1	7.4	24	12	170	11
Peat	713	7.2	6.8	7.7	14	12	15	46	40	54	33
Principal	5439	27	26	28	46	45	47	99	96	100	88

Vanadium (mg/kg)	n	50th	50th LCL	50th UCL	75th	75th LCL	75th UCL	95th	95th LCL	95th UCL	ULBL
Basalt	1803	180	180	180	220	220	230	280	270	280	350
Mournes	73	16	14	17	23	19	27	70	39	170	46
Peat	713	27	25	29	49	46	53	120	110	130	130
Principal	5439	84	83	86	110	110	110	140	140	140	170